From:	POULSEN Mike
To:	Burt Shephard/R10/USEPA/US@EPA
Cc:	Elizabeth Allen/R10/USEPA/US@EPA; Eric Blischke/R10/USEPA/US@EPA; Chip Humphrey/R10/USEPA/US@EPA; PETERSON Jenn L; jay.field@noaa.gov; mesl@shaw.ca; Joe Goulet/R10/USEPA/US@EPA
Subject:	RE: First draft of reliability statistics Powerpoint presentation for SETAC next week
Date:	11/02/2010 02:34 PM

Burt -

This is certainly an ambitious presentation. I have some suggestions for taking out a few slides, but then I add other slides, so I doubt that I'm helping to shorten the talk. I find it difficult to cut further because otherwise it would interfere with the logical flow of your presentation. Anyway, here are my thoughts.

- Mike

Slide 4 - I would drop this slide. You are at the end of the Portland Harbor session, so I think most people will have a good enough idea of the site. And you need to save time.

Slide 6 - Consider moving the prevalence discussion (slides 6 and 7) until after the presentation of reliability, but I don't feel strongly about this. I would shorten the values to two significant digits for clarity. Drop the text at the bottom. It likely won't be legible, and you can just make the statement orally.

Slide 7 - Same thoughts: cut the % values to 2 digits, and drop the text at the bottom.

Slide 8 - This will be very hard to read and understand. Drop the text at the bottom, and see if that helps. You might try making the A-B-C-D bins much bigger because explaining what they mean is the main point.

Slide 9 - Think about splitting this into two slides for better legibility. Remove text at bottom. You can shorten each bullet a little by dropping some words and saying, for example, "fraction of data correctly predicted". In BERA/SQG comments, EPA questioned the false negative goal of 0.20, so think about whether you want to provide the goal here.

Slide 11 - I would drop most of the text, and just include "SQB is too low", and "SQB is too high". The rest you can say.

Slide 12 - I would drop the text at the bottom (continuing my theme of removing text in small font that can't be read, and instead presenting the details orally).

Slides 13 & 14 - These are nice slides, but I don't think you have time to discuss them. Sadly, I suggest you drop them.

Slide 15 - This is very hard to read.

Slide 16 - The first bullet is a point where we may not completely agree, although the term "many" may cover it. As we discussed in the past, I consider false negative and false positive rates important for decision making. FN rates depend only on the hit dataset, without consideration of prevalence. And FP rates depend only on the no-hit dataset. Previously, I think you said we are both right, or that we agreed to disagree, or something. At any rate, I agree that overall reliability is dependent on prevalence, which is why I don't consider it a good measure of reliability.

Slide 18 - Try increasing the font of the sub-bullets.

Slide 19 - Drop the tiny print at the bottom.

Slide 20 - I think you should split this into two slides (1,2 and 3,4,5) for legibility.

Slides 21 & 22 - I don't think you have time to get into these slides. I suggest you drop them.

Slide 23 - The values for false negative rate and sensitivity (= 1 - FN) are switched (the false negative rates should be low). Highlight the best performing chance agreement. I suggest rounding each value to two digits to improve legibility.

I have not checked these results. Is this for a particular bioassay endpoint or combined? I'm a little surprised the FN rates for PEL and FPM values are the same. (The same for PNHR). NMI values are essentially the same for all three methods. So should we conclude that the reliabilities of the methods are the same, even though the SQBs vary by an order of magnitude? Looking at the FN rates, I would say No.

Slide 24 - I would split this into two slides for legibility.

Slide 25 - I would drop 5a), and just state this orally if you want. Regarding bullet 6, see my comment on slide 16. I'm not sure I agree with this statement. I do not think we should rely primarily on statistics such as kappa or NMI. However, because I don't see that false negative rates are affected by prevalence, perhaps I can agree. I would look at FN and FP positive rates, as well as kappa and NMI, in evaluating reliability.

-----Original Message-----

From: Shephard.Burt@epamail.epa.gov
[mailto:Shephard.Burt@epamail.epa.gov]
Sent: Tuesday, November 02, 2010 10:02 AM
To: POULSEN Mike; PETERSON Jenn L; jay.field@noaa.gov; mesl@shaw.ca; Goulet.Joe@epamail.epa.gov
Cc: Allen.Elizabeth@epamail.epa.gov; Blischke.Eric@epamail.epa.gov; Humphrey.Chip@epamail.epa.gov
Subject: First draft of reliabillity statistics Powerpoint presentation for SETAC next week

Attached is the first cut at a reliability statistics presentation to evaluate sediment toxicity predictive models and benchmarks for SETAC next week (Wednesday afternoon I believe). Any and all comments and suggestions are welcome, particularly anything that would shorten the talk (I have a couple of ideas here, such as eliminating one or two maps, and the table of sediment chemical concentration means and ranges). The poster presentation that goes along with this is largely the definitions, formulas and interpretation of the various reliability statistics. I wouldn't be able to get through all of them in a 20 minute talk in any event, so I didn't even try to present that in the Powerpoint. Instead, its a more general discussion of the problems of interpreting reliability in data sets with a low proportion of samples eliciting toxicity.

Best regards,

Burt Shephard
Risk Evaluation Unit
Office of Environmental Assessment (OEA-095)
U.S. Environmental Protection Agency, Region 10
1200 6th Avenue
Seattle, WA 98101

Telephone: (206) 553-6359
Fax: (206) 553-0119

e-mail: Shephard.Burt@epa.gov

"If your experiment needs statistics to analyze the results, then you ought to have done a better experiment"
- Ernest Rutherford

(See attached file: Shephard SETAC 2010 Portland Harbor platform 568.pptx)